UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

Rio Tinto plc,

Plaintiff,

-against-

Vale S.A., Benjamin Steinmetz, BSG Resources
Limited, VBG–Vale BSGR Limited aka BSG
Resources (Guinea) Ltd. aka BSG Resources
Guinée Ltd., BSG Resources Guinée SARL aka
BSG Resources (Guinea) SARL aka VBG-Vale
BSGR, Frederic Cilins, Mamadie Touré, and
Mahmoud Thiam,

Defendants.

---

14 Civ. 3042 (RMB) (AJP)

UNITED STATES DISTRICT COURT SONY
DOCUMENT
ELECTRONICALLY FILED
DOC#
DATE FILED: 9/8/15

STIPULATION AND ORDER RE: REVISED
VALIDATION AND AUDIT PROTOCOLS FOR
THE USE OF PREDICTIVE CODING IN DISCOVERY

WHEREAS, Plaintiff Rio Tinto and Defendant Vale (collectively, the "Parties"

and each, a "Party") agreed to the use of predictive coding for the search, review, and

production of documents in the above-captioned litigation, and entered a stipulation

memorializing their agreement (Dkt. No. 206) (hereinafter, the "Predictive Coding

Protocol");

WHEREAS, during the course of the Parties' Predictive Coding Processes,[1]

disputes arose relating, in particular, to the validation process and the adequacy of the

Parties' document productions pursuant to the Predictive Coding Processes;

---

[1] Unless otherwise indicated, all defined terms have the meanings assigned to them in the
Parties' Predictive Coding Protocol.

WHEREAS, on July 15, 2015, the Court appointed a Special Master "to resolve such disputes that may arise in relation to the use of technology assisted review (aka predictive coding) by the parties. . . ." (Dkt. No. 301);

WHEREAS, on July 29 and August 4, 5, 10, and 12, 2015, the Parties jointly and individually submitted information and arguments concerning their disputes, the Predictive Coding Processes, and the adequacy of the productions pursuant to the Predictive Coding Processes;

WHEREAS, on August 19, 2015, the Special Master met with the parties to resolve their disputes concerning the Predictive Coding Processes and the adequacy of the productions pursuant to the Predictive Coding Processes;

IT IS HEREBY STIPULATED AND AGREED, by and between the undersigned, as attorneys of record for the Parties, as follows:

1.   Rio Tinto disclosed that it used a "broad, concept-based approach" to responsiveness during the predictive coding training process, and a "narrower, request-for-production-based approach" to responsiveness during the (post-predictive coding) manual review process; therefore, a new Control Set is needed to estimate prevalence, and the Precision and Recall of Rio Tinto's Predictive Coding Process.

2.   The parties had previously agreed that following the supplemental training agreed to by the Parties at the meet and confer that preceded the June 22, 2015 status conference, and adopted by the Court (June 22, 2015 Hr'g Tr. at 12:16-13:24), and after production was complete, they would repeat the validation process set forth in Section 4(f) of the Parties' Predictive Coding Protocol. Rio Tinto agrees that, in lieu of repeating this validation process, it shall undertake the Revised Validation Protocol set forth below:

(a)   After production is complete, Rio Tinto's predictive coding provider shall draw a random sample of 2,399 documents from the Document Universe (the "New Control Set") and shall place them in a folder that contains no information about how the documents were previously coded or whether they were previously produced or withheld, and the basis therefor.

(b)   Rio Tinto shall review and code the New Control Set and shall provide a certification under oath that the reviewer(s) of its New Control Set had no knowledge of how the documents were previously coded or whether they were produced or withheld, and the basis therefor.

(c)   After the review and coding of the New Control Set is complete, and, in any event, no later than September 8, 2015, per the process outlined for the original Control Set in the Parties' Predictive Coding Protocol, Rio Tinto shall disclose the results of the review of the New Control Set to Vale, including the number of documents that were coded for each of the Responsiveness Categories during the review of the New Control Set. Rio Tinto shall also produce all non-privileged documents reviewed in the New Control Set, and for each document disclose the Responsiveness Category or Categories to which it belongs.

(d)   Vale shall raise any disputes regarding how the documents in the New Control Set were coded for each of the Responsiveness Categories within five (5) business days of the production. The Parties agree to meet and confer in good faith over any such disputes. All non-responsive documents produced from the New Control Set shall be deemed "Highly Confidential" under the terms of the Stipulated Protective Order (Dkt. No. 81), shall be used only for the purpose of evaluating the accuracy of the

document coding, and shall be promptly returned or destroyed after review by Vale and the resolution of any disputes. Should the Parties be unable to resolve their disputes on any issues stemming from the Revised Validation Protocol, they shall promptly submit those issues to the Special Master, or as appropriate, to the Magistrate Judge, for resolution.

(e)   Within five (5) business days of final agreement on the coding of the New Control Set, Rio Tinto shall complete and provide Vale and the Special Master with TAR Worksheet A (attached hereto as Exhibit A), and indicate for each document in the New Control Set whether it was: (i) produced as responsive in its own right; (ii) produced as a family member of a responsive document; (iii) withheld as privileged; (iv) withheld as non-responsive; (v) not identified as potentially responsive by the Predictive Coding Process; or (vi) not produced for any other reason(s), which reason(s) shall be provided. Rio Tinto shall also calculate an estimate of prevalence, Precision, and Recall, at the 95% confidence level, using the JavaStat Exact Binomial Confidence Interval Calculator found at statspages.org/confint.html, including the applicable confidence intervals.

(f)   If the Parties are unable to reach agreement on whether the Revised Validation Protocol confirms that Rio Tinto has made an adequate production, the Special Master shall determine (taking into account all available evidence, including the results of the Audit Protocol outlined below) whether Rio Tinto's production is adequate, inadequate, or indeterminate because the confidence interval is too wide, and, if inadequate or indeterminate, what further action needs to be taken by Rio Tinto.

3.   Vale agrees to repeat the validation process set forth in Section 4(f) of the Parties' Predictive Coding Protocol, and to disclose the results to Rio Tinto and the Special Master. If the Parties are unable to reach agreement on whether the validation process confirms that Vale has made an adequate production, Rio Tinto may raise its concerns with the Special Master, Vale may respond to those concerns, and the Special Master shall determine whether Rio Tinto has shown a material problem or deficiency, and if so, what further action needs to be taken by Vale.

4.   Both Rio Tinto and Vale have raised issues and concerns regarding the other's Predictive Coding Process and the productions pursuant thereto. In particular, one or both parties have questioned the use of search terms prior to predictive coding, the adequacy of the manual review after predictive coding, and/or the failure of the other Party to produce documents in response to certain requests for production.

5.   The validation process set forth in the Parties' Predictive Coding Protocol focuses on the overall Recall achieved as a result of the predictive coding training process; it does not address the keyword culling process before predictive coding, the manual review process after predictive coding, or individual facets of responsiveness.[2] Accordingly, to address these issues more directly, the Parties shall undertake the Audit Protocol set forth below:

(a)   Each Party shall be allotted a total of 2,400 documents for the purposes of its audit (the "Audit Sample"). The Requesting Party may use its Audit

---

[2] These issues are discussed in Maura R. Grossman and Gordon V. Cormack, *Comments on "The Implications of Rule 26(g) on the Use of Technology-Assisted Review,"* 7 Fed. Courts L. Rev. 285, 300-10 (2014), *available at* http://www.fclr.org/fclr/articles/pdf/comments-implications-rule26g-tar-62314.pdf. There is nothing about these issues that is unique to technology-assisted review; they can arise in any review process.

Sample to test whatever aspect(s) of the Responding Party's search, review, and production process it wishes, provided that the Responding Party's predictive coding provider is capable of undertaking the request without undue cost or burden. For example, and not by way of limitation, Rio Tinto may request that a random sample of documents be drawn from the documents that did not "hit" on any of Vale's search terms; either party may request that a random sample of documents be drawn from the documents coded as "non-responsive" by the Predictive Coding Software, and therefore not subject to subsequent manual review; either party may request that a random sample of documents be drawn from the documents coded as "responsive" by the Predictive Coding Software, but as "non-responsive" during the subsequent manual review; either party may supply one or more search criteria (regardless of whether such search criteria previously have been provided) and request that a random sample of the "hits" (or the top-ranked "hits") be drawn; and so on. The Audit Sample may be divided into as many categories (each, an "Audit Category") as the Requesting Party would like, provided that the total Audit Sample shall not exceed 2,400 documents.

(b)   The Parties shall exchange (and shall provide the Special Master with) their request for their Audit Samples by no later than September 14, 2015, which shall include: (i) a description of the method by which the documents should be selected (*e.g.*, keyword "hits," "clustering," random sampling, etc.); (ii) the population from which the documents should be drawn (*e.g.*, documents that were not keyword "hits," documents that were identified as likely non-responsive by the Predictive Coding Process, documents that were marked as "non-responsive" by the manual reviewers, etc.); and (iii) the number of documents to be included in that Audit Category. The Parties

shall meet and confer in good faith to resolve any questions or technical issues raised by the request for the Audit Samples.

(c)   The Responding Party shall have its predictive coding provider draw the Audit Sample as specified. If the same document is selected for more than one Audit Category, it shall be reviewed only once, but it shall be counted towards any Audit Category for which it is selected. The predictive coding provider shall maintain a separate record of which documents were selected for which Audit Categories, and shall place the Audit Sample in a folder that contains no information about the Audit Categories, how the documents were previously coded or whether they were produced or withheld, and the basis therefor.

(d)   The Responding Party shall review and code the Audit Sample and shall provide a certification under oath that the reviewer(s) of its Audit Sample had no knowledge of the Audit Categories, how the documents were previously coded or whether they were produced or withheld, and the basis therefor. The Parties have agreed that the Responding Party shall also provide all non-privileged documents in the Audit Sample, which production shall be made within five (5) business days of receipt of the Audit Sample request.[3]

(e)   The Requesting Party shall raise any disputes regarding how the documents in the Audit Sample were coded for each of the Responsiveness Categories within five (5) business days of its production. The Parties agree to meet and confer in good faith over any such disputes. All non-responsive documents produced from the Audit Sample shall be deemed "Highly Confidential" under the terms of the Stipulated

---

[3] It should be noted that the mutual disclosure of non-relevant documents was requested by the Parties, not the Special Master.

Protective Order (Dkt. No. 81), shall be used only for the purpose of evaluating the accuracy of the document coding, and shall be promptly returned or destroyed after review by the Requesting Party and the resolution of any disputes. Should the Parties be unable to resolve their disputes on any issues stemming from the Audit Protocol, they shall promptly submit those issues to the Special Master, or as appropriate, to the Magistrate Judge, for resolution.
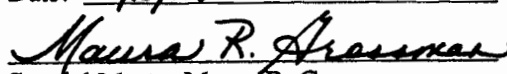
(f)   Within five (5) business days of final agreement on the coding of the Audit Sample, the Responding Party shall complete and provide the Requesting Party and the Special Master with TAR Worksheet B (attached hereto as Exhibit B), and indicate for each document in the Audit Sample, the Audit Category or Categories to which it belongs, and whether it was: (i) produced as responsive in its own right; (ii) produced as a family member of a responsive document; (iii) withheld as privileged; (iv) withheld as non-responsive; (v) not identified as potentially responsive by the Predictive Coding Process; or (vi) not produced for any other reason(s), which reasons(s) shall be provided.
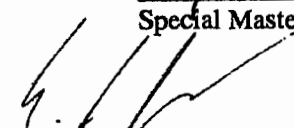
(g)   If the Parties are unable to reach agreement on whether the Audit Protocol confirms that the Responding Party has made an adequate production with respect to any particular Audit Category, or if the Parties agree that the production is inadequate with respect to a particular Audit Category, but cannot agree on the appropriate remediation procedure, the Special Master shall determine (taking into account all available evidence) whether the production is adequate, inadequate, or indeterminate, and, if inadequate or indeterminate, what further action needs to be taken by the Responding Party.

6.   This Stipulation and Order resolves all disputes between the Parties

concerning the Predictive Coding Processes and the productions pursuant to the

Predictive Coding Processes, other than those that may arise as a result of the Revised

Validation Protocol outlined in (2.) above for Rio Tinto, the validation process outlined

in (3.) above for Vale, and the Audit Protocol outlined in (5.) above for both Parties. The

Parties reserve all rights to challenge the adequacy of the opposing Party's production (or

any facet of that production) based on the outcome of the Revised Validation and Audit
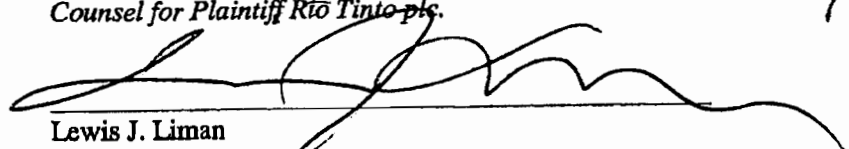
Protocols outlined herein.

IT IS SO ORDERED

Date: *9/2/15*

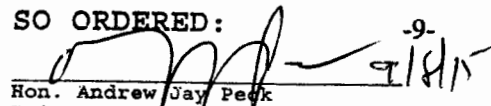Special Master Maura R. Grossman

Eric C. Lyttle
Meghan McCaffrey
QUINN EMANUEL URQUHART & SULLIVAN, LLP
777 6th Street NW, 11th floor
Washington, DC 20001
ericlyttle@quinnemanuel.com
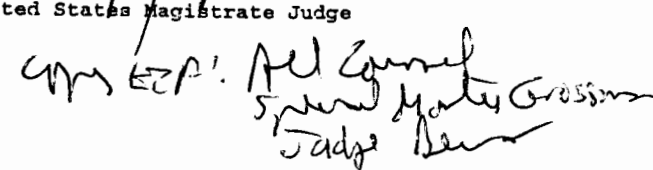mmccaffrey@quinnemanuel.com

Counsel for Plaintiff Rio Tinto plc.

Lewis J. Liman
Scott Reents
CLEARY GOTTLIEB STEEN & HAMILTON LLP
One Liberty Plaza
New York, NY 10006
lliman@cgsh.com
sreents@cgsh.com

Counsel for Defendant Vale S.A.

SO ORDERED:

-9-

Hon. Andrew Jay Peck
United States Magistrate Judge

EXHIBIT A

TAR WORKSHEET A

1. How many documents were in the document collection ("Document Collection") from which the New Control Set was derived? *For the purposes of this TAR Worksheet, a "document" is defined as a single loose file, the body of an email (often referred to as a "parent" email), or a single attachment to an email.* The "New Control Set" is the set of 2,399 documents drawn at random from the Document Collection, according to the Revised Validation Protocol, with the final responsiveness coding determinations as agreed to by the parties (the "Final Responsiveness Coding").

   Answer: Click here to enter text.

2. How many documents in the New Control Set had a Final Responsiveness Coding as follows?

   (a) Responsive: Click here to enter text.

   (b) Non-Responsive: Click here to enter text.

3. How many documents in the Document Collection were:

   (a) Reviewed for production as a result of the Predictive Coding Process (including documents classified by the Predictive Coding Software, documents that could not be classified by the Predictive Coding Software, documents identified through targeted searches, etc.)?

      Answer: Click here to enter text.

   (b) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process?

      Answer: Click here to enter text.

   (c) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process [as per your response to question 3(b)], but withheld as Privileged?

      Answer: Click here to enter text.

   (d) Reviewed for production and coded as Non-Responsive as a result of the Predictive Coding Process?

      Answer: Click here to enter text.

4. How many documents in the New Control Set were:

(a) Reviewed for production as a result of the Predictive Coding Process?

Answer: Click here to enter text.

(b) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process?

Answer: Click here to enter text.

(c) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process [as per your response to question 4(b)], but withheld as Privileged?

Answer: Click here to enter text.

(d) Reviewed for production and coded as Non-Responsive as a result of the Predictive Coding Process?

Answer: Click here to enter text.

(e) Reviewed for production and coded as Non-Responsive as a result of the Predictive Coding Process [as per your response to question 4(d)], but withheld as Privileged?

Answer: Click here to enter text.

5. How many documents with a Final Responsiveness Coding of Responsive in the New Control Set [as per your response to 2(a)] were:

(a) Reviewed for production as a result of the Predictive Coding Process?

Answer: Click here to enter text.

(b) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process?

Answer: Click here to enter text.

(c) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process [as per your response to question 5(b)], but withheld as Privileged?

Answer: Click here to enter text.

(d) Reviewed for production and coded as Non-Responsive as a result of the Predictive Coding Process?

Answer: Click here to enter text.

(e) Reviewed for production and coded as Non-Responsive as a result of the Predictive Coding Process [as per your response to question 5(d)], but withheld as Privileged?

Answer: Click here to enter text.

6. Please compute the answers to questions (a) through (d) below using the JavaStat Exact Binomial Confidence Interval Calculator found at statpages.org/confint.html, and the answer to question (e) using an ordinary calculator.

    (a) Prevalence: [x = your answer to question 2(a) and N = 2399]

        Proportion: Click here to enter text.

        Exact Confidence Interval: Click here to enter text. to Click here to enter text.

    (b) Recall: [x = your answer to question 5(b) and N = your answer to question 2(a)]

        Proportion: Click here to enter text.

        Exact Confidence Interval: Click here to enter text. to Click here to enter text.

    (c) Precision: [x = your answer to question 5(b) and N = your answer to question 4(b)]

        Proportion: Click here to enter text.

        Exact Confidence Interval: Click here to enter text. to Click here to enter text.

    (d) Expected proportion of Document Collection coded as Responsive as a result of the Predictive Coding Process, irrespective of whether produced or withheld as Privileged: [x = your answer to question 4(b) and N = 2399]

        Proportion: Click here to enter text.

        Exact Confidence Interval: Click here to enter text. to Click here to enter text.

    (e) Actual proportion of the Document Collection coded as Responsive as a result of the Predictive Coding Process, irrespective of whether produced or withheld as Privileged: [divide your answer to question 3(b) by your answer to question 1]

        Proportion: Click here to enter text.

        [There is no confidence interval, as this is a precise number, not an estimate.]

7. In addition to this TAR Worksheet, have you supplied an electronic file containing the information for each document in the New Control Set required by Section (e) of the Revised Validation Protocol?

Answer: Yes ☐

I swear under penalty of perjury that (i) the reviewer(s) who coded the documents in the New Control set had no knowledge of how the documents contained therein were previously coded or whether they were produced or withheld, and the basis therefor, and (ii) that the answers provided in this TAR Worksheet are complete and correct to the best of my knowledge, information, and belief.

Signature: ×

Name: Click here to enter text.

Title and Name of Organization: Click here to enter text.

Date: Click here to enter text.

EXHIBIT B

TAR WORKSHEET B

Please complete a copy of this TAR Worksheet for each Audit Category in the Audit Request. For the purposes of this TAR Worksheet, a "document" is defined as a single loose file, the body of an email (often referred to as a "parent" email), or a single attachment to an email.

1. Indicate the name of the Audit Category.

   Answer: Click here to enter text.

2. How many of the 2,400 documents in the Audit Sample were alloted to this Audit Category? These documents, along with the final coding determinations agreed to by the parties, are hereinafter referred to as the "Audit Subsample."

   Answer: Click here to enter text.

3. Indicate the number of documents in the Audit Subsample that were coded as:

   (a) Responsive: Click here to enter text.

   (b) Non-Responsive: Click here to enter text.

4. How many documents that were coded as Responsive in the Audit Subsample [as per your response to 3(a)] were:

   (a) Reviewed for production as a result of the Predictive Coding Process (including documents classified by the Predictive Coding Software, documents that could not be classified by the Predictive Coding Software, documents identified through targeted searches, etc.)?

   Answer: Click here to enter text.

   (b) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process?

   Answer: Click here to enter text.

   (c) Reviewed for production and coded as Responsive as a result of the Predictive Coding Process [as per your response to question 4(b)], but withheld as Privileged?

   Answer: Click here to enter text.

   (d) Reviewed for production and coded as Non-Responsive as a result of the Predictive Coding Process?

Answer: Click here to enter text.

(e) Reviewed for production and coded as Non-Responsive as a result of the Predictive Coding Process [as per your response to question 4(d)], but withheld as Privileged?

Answer: Click here to enter text.

5. Please compute the answers to the question below using the JavaStat Exact Binomial Confidence Interval Calculator found at statpages.org/confint.html.

Recall: [x = your answer to question 4(b) and N = your answer to question 3(a)]

Proportion: Click here to enter text.

Exact Confidence Interval: Click here to enter text. to Click here to enter text.

Please provide one executed copy of this declaration with the set of your completed copies of TAR Worksheet B for each Audit Category.

In addition to this TAR Worksheet, I have supplied an electronic file containing the information for each document in the Audit Sample required by Section (f) of the Audit Protocol.

I swear under penalty of perjury that (i) the reviewer(s) who coded the documents in the Audit Sample had no knowledge of the Audit Categories, how the documents contained therein were previously coded or whether they were produced or withheld, and the basis therfor, and (ii) that the answers provided in this TAR Worksheet are complete and correct to the best of my knowledge, information, and belief.

Signature: _×_____

Name: Click here to enter text.

Title and Name of Organization: Click here to enter text.

Date: Click here to enter text.